



Experimentation and social interventions: a forgotten but important history

Ann Oakley

BMJ 1998;317;1239-1242

Updated information and services can be found at:
<http://bmj.com/cgi/content/full/317/7167/1239>

These include:

References

10 online articles that cite this article can be accessed at:
<http://bmj.com/cgi/content/full/317/7167/1239#otherarticles>

Rapid responses

One rapid response has been posted to this article, which you can access for free at:
<http://bmj.com/cgi/content/full/317/7167/1239#responses>

You can respond to this article at:
<http://bmj.com/cgi/eletter-submit/317/7167/1239>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections

Articles on similar topics can be found in the following collections

[Sociology](#) (314 articles)

[Randomized Controlled Trials: descriptions](#) (305 articles)

Notes

To order reprints of this article go to:
<http://www.bmjournals.com/cgi/reprintform>

To subscribe to *BMJ* go to:
<http://bmj.bmjournals.com/subscriptions/subscribe.shtml>

will generate an inordinate number of telephone calls and mountains of correspondence, mailshots, newsletters, faxes, and emails. It is crucial that the secretary is pleasant and efficient on the telephone and is able to prioritise his or her work.

Large trials must not produce huge amounts of work for the body of clinicians willing to participate. Consequently, this means a great deal of work for the team of the coordinating centre. Efficiency is paramount. The systems and procedures within the coordinating centre must be simple and focused on the task. Even simply date stamping every piece of paper that comes into the trial office has a purpose, and its value cannot be overemphasised. A data manager working on the long term follow up of a trial should be focused on that particular task and not be distracted by other needs of the trial office. Members of the trial team need to be familiar with all aspects of the trial but should have an in depth knowledge of their own area of work, not only for the trial but for their job satisfaction and career development.

A group of clinicians come together because they want to contribute to the science that forms the basis of their everyday practice. The trial team needs to nurture the collaborative group by helping with the boring

mundane aspects of a trial. Nothing should be too much trouble for the trial team: the team should not wait for collaborators to ask for help but should be proactive and make life easier for them. Most importantly, the team should motivate and instil a sense of ownership among the collaborators. This should lead to swift recruitment and good quality data.

A trial cannot be managed by a committee. Although a large randomised trial needs a steering committee to give direction on policy, oversee the professional conduct of the group, and give an independent opinion on management matters, it is not a substitute for trial management. As trial coordinators or managers, we should seek to increase the knowledge base from which trials operate. In order to do this we must value our trial management skills, pass on experience, and develop clinical trial management in a meaningful and professional way.

- 1 *The Cochrane Library* [database on disk and CD ROM]. Cochrane Collaboration. Oxford: Update Software, 1997. Updated quarterly.
- 2 UK Medical Research Council. *Good clinical practice guidelines for clinical trials*. London: UK Medical Research Council, 1998.
- 3 Warlow CP. How to do it: Organise a multicentre trial. *BMJ* 1990;300:180-3.

(Accepted 5 October 1998)

Experimentation and social interventions: a forgotten but important history

Ann Oakley

The research design of the randomised controlled trial is primarily associated today with medicine. It tends either to be ignored or regarded with suspicion by many in such disciplines as health promotion, public policy, social welfare, criminal justice, and education. However, all professional interventions in people's lives are subject to essentially the same questions about acceptability and effectiveness. As the social reformers Sidney and Beatrice Webb pointed out in 1932, there is far more experimentation going on in "the world sociological laboratory in which we all live" than in any other kind of laboratory, but most of this social experimentation is "wrapped in secrecy" and thus yields "nothing to science."¹

The Webbs argued for a more "scientific" social policy, with social scientists being trained in experimental methods and evaluations of social interventions being carried out by independent investigators. They were apparently unaware that a strong tradition in experimental sociology had already been established, mainly in the United States. This was a precursor to a period between the early 1960s and the late 1980s when randomised controlled trials became the ideal for American evaluators assessing a wide range of public policy interventions. This history is conveniently overlooked by those who contend that randomised controlled trials have no place in evaluating social interventions. It shows clearly that prospective experimental studies with random allocation to generate one or more control groups is perfectly possible in social settings. Notably, too, the history of

Summary points

Many social scientists argue that randomised controlled trials are inappropriate for evaluating social interventions, but they ignore a considerable history, mainly in the United States, of the use of randomised controlled trials to assess different approaches to public policy and health promotion

A tradition of experimental sociology was well established by the 1930s, built on the early use of controlled experiments in psychology and education

From the early 1960s to early 1980s randomised experiments were considered the optimal design for evaluating public policy interventions in the United States, and major evaluations using this design were carried out

This approach became less popular as policy makers reacted negatively to evidence of "near zero" effects

Lessons to be learnt about implementing randomised controlled trials in real life settings include the difficulty of assessing complex multi-level interventions and the challenge of integrating qualitative data

Social Science Research Unit,
University of London Institute of Education, London WC1H 0NS
Ann Oakley,
professor
a.oakley@ioe.ac.uk

BMJ 1998;317:1239-42

experimentation in social science predates that in medicine in certain key respects.

A short history of control groups

The original meaning of “control” is “check”—the word comes from “counter-roll,” a duplicate register or account made to verify an official account.² The term “control” entered scientific language in the 1870s in the sense of a standard of comparison used to check inferences deduced from an experiment. The main use of the term was in experimental psychology.³

In 1901 the American educationalists Thorndike and Woodworth identified the need for a control group in their experiments on the use of training to improve mental function.⁴ A series of experiments with schoolchildren that addressed questions about the transferability of memory skills from one subject to another, reported by Winch in 1908,⁵ were among the first to use the design of pretest, intervention, post-test in the experimental group and pretest, nothing, post-test in the control group. These educational and psychology researchers invented randomised assignment to experimental treatments and Latin square designs independently of, and considerably earlier than, R A Fisher’s work at the Rothamsted Agricultural Research Station.⁶ The psychologist C S Peirce introduced both the idea of randomisation and that of “blindness” into psychology experiments in the 1880s.⁷

Selection of experimental and control subjects by means of the principle of chance is described in McCall’s *How to Experiment in Education*, published in 1923: “Representativeness [of research subjects] can be secured by making a chance selection from the total group, or a chance selection from a chance portion of the total group Just as representativeness can be secured by the method of chance, so equivalence may be secured by chance One method of equating by chance is to mix the names of the subjects to be used. Half may be drawn at random. This half will constitute one group while the other half will constitute the other group.”⁸ McCall’s book also describes the Latin square design under the name of the “rotation experiment”; this had been used in educational experiments as early as 1916.⁹

The major impetus driving these new approaches to assessing effectiveness was not the desire to imitate natural science, but, rather, to respond to an uneasiness within the research community of educational psychology about the inability of existing evaluation methods to rule out plausible rival hypotheses. Similar methodological developments were occurring in other spheres. For example, in 1924-5 an experiment using a mail campaign to increase electoral turnout was carried out in Chicago, in which housing precincts were assigned either to receive individual mail appeals or not.¹⁰ This experiment followed earlier research which had suggested that the strength of local party organisation was the main factor distinguishing voters from non-voters, but the research design used in the first study had made it impossible to have confidence in this finding. Thus, in the social field as well as later in medicine, the advantages of prospective experimental studies with randomly chosen controls were seen to offer an important solution to the problem of linking intervention with outcomes.

Experimental sociology

Two other American social scientists, Ernest Greenwood at Columbia University and F Stuart Chapin at the University of Minnesota, pioneered the application of experimental methods to the study of social problems in the early decades of the 20th century. Chapin first wrote on this theme in 1917; his *Experimental Designs in Sociological Research*, published in 1947, details nine experimental studies carried out by his research team and a number undertaken elsewhere covering such topics as rural health education, the social effects of public housing, recreation programmes for “delinquent” boys, and the effects of student participation in extracurricular activities.¹¹ Chapin was particularly interested in reviewing the use of experimental research designs in “the normal community situation” because of the objection, voiced at the time, that experimental studies could only be done in “laboratory” settings.

Ernest Greenwood’s *Experimental Sociology*, published in 1945, outlined the theoretical rationale for applying experimental methods to social issues.¹² He defined an experiment as “the proof of a causal hypothesis through the study of two controlled contrasting situations,” recommended the use of case studies as a prelude to experimental research, and supported Fisher’s strategy of randomisation as the best way of securing equivalent study groups. Chapin’s and Greenwood’s interest in experimental research designs was stimulated by the social reform concerns of the Depression, and informed by a desire to establish the most effective methods of improving people’s lives. Their work was part of a general move in the United States to make social science more experimental; by 1931 at least 26 universities there were offering courses in experimental sociology.¹³

A golden age of evaluation

Donald Campbell and Julian Stanley’s *Experimental and Quasi-experimental Designs for Research* published in 1966¹⁴ is to social research what Fisher’s *Design of Experiments* (1935) is to medical research. Campbell’s paper “Reforms as experiments” established an explicit link between social reform and the use of rigorous experimental design.¹⁵ His complaint that the randomised control group design had not often been used in the social arena prompted another American experimentalist, Robert Boruch, to publish a bibliography of these in 1974.¹⁶ This listed 83 “randomised field experiments” in such areas as criminal justice, legal policy, social welfare, education, mass communications, and mental health. A revised version of the bibliography produced four years later updated the total in these areas to 245.¹⁷

This period in the United States has been nicknamed the “golden age of evaluation.”¹⁸ It was one in which there was an enormous burst of activity in applying the randomised controlled trial design to the evaluation of public policy. The table shows nine of the major evaluations of broadly based social programmes initiated between the 1960s and early 1980s. Four of the studies were of income maintenance experiments,¹⁹⁻²³ one focused on an experimental housing allowance scheme,^{24 25} two examined pro-

Examples of controlled trials of social programmes carried out in the United States during 1968-91

Trial	Years	Aim	Design	Outcomes assessed
Income maintenance experiments				
New Jersey-Pennsylvania negative income tax experiment (see Ferber and Hirsch, ¹⁹ Rossi and Lyall ²⁰)	1968-72	To study effects on work incentives of negative income tax	Random allocation of 1216 low income families to 8 intervention and 1 control groups	Participation in labour force; consumption expenditure; health and family behaviour; school attendance
Rural negative income tax experiment (see Ferber and Hirsch, ¹⁹ Maynard ²¹)	1970-2	To replicate above experiment in poor rural areas with non-intact families with female or male heads	Stratified random allocation of 809 low income families to 5 intervention and 1 control groups	Participation in labour force; consumption expenditure; health and family behaviour; school attendance
Gary income maintenance experiment (see Ferber and Hirsch, ¹⁹ Kehrer, ²² Kehrer and Wolin ²³)	1971-4	To study effects on participation in labour force and other family behaviours of different levels and forms of income maintenance, day care subsidies, and information and referral services	Stratified random allocation of 1799 low income, single parent families	Participation in labour force; consumption expenditure; health and family behaviour; school attendance; social and psychological attitudes
Denver-Seattle income maintenance experiments (see Ferber and Hirsch, ¹⁹ Rossi and Lyall ²⁰)	1970-91	To study effects on participation in labour force and other household behaviours of different levels and forms of income maintenance, job counselling, and training subsidies	Stratified random allocation of 2042 families allocated to 84 experimental "cells" with different combinations of support levels, tax rates, etc, and 1 control group	Participation in labour force; consumption expenditure; health and family behaviour; school attendance
Housing allowances				
Experimental housing allowance program (demand experiment) (see Bradbury and Downs, ²⁴ Friedman and Weinberg ²⁵)	1978-80	To study effects on households' housing behaviour of different forms of housing allowances and estimate of cost effectiveness	Stratified random allocation of 2241 low income households to 17 intervention groups with different housing allowance formulae and 2 control groups	Quality of housing; housing consumption behaviour; mobility
Supporting workers programmes				
Supported work program (see Ferber and Hirsch ¹⁹)	1975-8	To study effects and costs of supported work environment for disadvantaged workers	Random allocation of 6616 disadvantaged workers to 1 intervention and 1 control group	Participation in labour force; hours worked; total earnings
Texas worker adjustment program (see Bloom ²⁶)	1984-5	To study effects and costs of combination of job search assistance and occupational skills training for displaced workers	Random allocation of 2259 hard to employ individuals by random numbers table to 2 intervention and 1 control groups on 1 site, and 1 intervention and 1 control groups on 2 sites	Earnings; unemployment; unemployment benefits
Penal experiments				
Living insurance for ex-prisoners (LIFE) (see Rossi et al ²⁷)	1971-4	To study effects on re-arrests and participation in labour force of different levels of post-release payment and job assistance schemes	Stratified random allocation of 432 released prisoners to 3 intervention groups (payments only, counselling and placement only, both combined) and 1 control group	Arrests and convictions by type of offence; participation in labour force; health and living arrangements
Transitional aid research project (TARP) (see Rossi et al ²⁷)	1975-7	To study effects on re-arrests and participation in labour force of different levels of post-release payment and job assistance schemes	Stratified random allocation of 3982 released prisoners to 4 intervention groups with combinations of different payment periods and tax rates (3 groups) and job placement services (1 group) plus 2 control groups	Arrests and convictions by type of offence; participation in labour force; health and living arrangements

grammes for supporting disadvantaged workers,^{19, 26} and two examined interventions for former prison inmates.²⁷ All the studies included one or more prospectively generated control groups, either by some method of random allocation or by matching. Supporting all this effort was a government mandate specifying that 1% of budgets for social programmes had to be spent on evaluation. There was widespread recognition that social services were in a mess while expenditure on them was rising exponentially; and, for a time at least, there was a consensus in policy circles that randomised controlled experiments provided the best way of assessing effectiveness.

Other evaluations (not shown in the table) carried out during this period included the Manhattan bail bond experiment with pre-trial release for prisoners,²⁸ the Rand Corporation's well known study of health insurance (several components of which used a randomised controlled trial design),²⁹ and studies of educational performance contracting.³⁰

The reasons why the use of randomised controlled trials in evaluating policy interventions has declined in attractiveness in the United States over the past 20 years are as interesting as those explaining its

acceptance in the first place. A primary one was disenchantment with the apparent ineffectiveness (sometimes seemingly damaging effects) of the interventions in some of the evaluations. Secondly, policy makers were often impatient with the length of time it took for evaluations of their favoured approaches to provide answers: this was particularly marked in the case of the income experiments. As Senator Moynihan appositely said, "The bringing of systematic inquiry to bear on social issues is not an easy thing. There is no guarantee of pleasant and simple answers, but if you make a commitment to an experimental mode it seems to me ... something larger is at stake when you begin to have to deal with the results."³¹

Conclusions

All claims to successful expertise need to tackle the issue of causal inference—how do people know that what they do works, and how can they reasonably demonstrate this to others? As Stanley noted in 1957, "Expert opinions, pooled judgements, brilliant intuitions, and shrewd hunches are frequently misleading."³² Among the reasons why randomised controlled trials

gained legitimacy in medicine was the realisation that the decisions of the medical profession need to be regulated.³³ The history of social experimentation indicates clearly that all the same issues have attended attempts to evaluate the impact of social interventions.

Experts in the social domain, like those in medicine, have resisted the notion that rigorous evaluation of their work is more likely to give reliable answers than their own individual preferences. When randomised controlled trials find that new "treatments" are no better than old ones, a retreat to other methods of evaluation is particularly likely, as though the prime task is not to identify whether anything works but to prove that something does.

The forgotten history of social experimentation also shows that, as in clinical research, implementing randomised controlled trials in real life settings commonly carries a number of hazards: low participation rates or high attrition, problems with "informed consent," unanticipated side effects of the intervention, a problematic relation between research and policy.

There are many lessons to be learnt from this experience about the challenges of randomised controlled trials, including the difficulty of establishing the effectiveness of complex multi-level interventions and the problem of integrating ethnographic or qualitative data. But, as Chapin wrote in 1931, "Experimental method in sociology does not mean interference with individual movement or freedom. It does not endanger life or limb or moral character."³⁴ On the contrary, what randomised controlled trials offer in the social domain is exactly what they promise to medicine: protection of the public from potentially damaging uncontrolled experimentation and a more rational knowledge about the benefits to be derived from professional intervention.

- 1 Webb S, Webb B. *Methods of social study*. London: Longmans Green 1932:224-5.
- 2 Boring EG. The nature and history of experimental control. *Am J Psychol* 1954;LXVII:573-89.
- 3 Murphy G, Murphy LB, Newcomb TM. *Experimental social psychology*. New York: Harper and Bros, 1937.
- 4 Thorndike EL, Woodworth RS. The influence of improvement in one mental function upon the efficiency of other functions. *Psychol Rev* 1901;8:247-61, 384-95, 553-64.
- 5 Winch WH. The transfer of improvement in memory in school-children. *Br J Psychol* 1908;2:284-93.

- 6 Campbell DT. *Methodology and epistemology for social science: selected papers*. Chicago: University of Chicago Press, 1988:321.
- 7 Stigler SM. *The history of statistics*. Cambridge, MA: Belknap Press, 1986:253.
- 8 McCall WA. *How to experiment in education*. New York: Macmillan, 1923:38-41.
- 9 Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin, 1966:2.
- 10 Gosnell HF. *Getting out the vote: an experiment in the stimulation of voting*. Westport, CT: Greenwood Press, 1977. (Originally published 1927.)
- 11 Chapin FS. *Experimental designs in sociological research*. New York: Harper and Row, 1947.
- 12 Greenwood E. *Experimental sociology*. New York: Octagon Books, 1976:72. (Originally published 1945.)
- 13 Brearley HC. Experimental sociology in the United States. *Social Forces* 1931;Dec:196-9.
- 14 Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin, 1966.
- 15 Campbell DT. Reforms as experiments. *Am Psychol* 1924;4:409-29.
- 16 Boruch RF. Bibliography: illustrative randomized field experiments for program planning and evaluation. *Evaluation* 1974;2:83-7.
- 17 Boruch RF. Randomized field experiments for program planning, development, and evaluation. *Eval Q* 1978;2(4):655-95.
- 18 Rossi PH, Wright JD. Evaluation research: an assessment. *Ann Rev Sociol* 1984;10:331-52.
- 19 Ferber R, Hirsch WZ. *Social experimentation and economic policy*. Cambridge: Cambridge University Press, 1982.
- 20 Rossi PH, Lyall KC. *Reforming public welfare*. New York: Russell Sage Foundation, 1976.
- 21 Maynard R. The effects of the rural income maintenance experiment on the school performance of children. *Am Econ Rev* 1977;67:370-5.
- 22 Kehrer KC. The Gary income maintenance experiment: summary of initial findings. In: Cook TD, DelRosario ML, Hennigan KM, Mark MM, Trochim WMK, eds. *Evaluation studies review annual*. No 3. Beverly Hills, CA: Sage, 1978:437-44.
- 23 Kehrer KC, Wolin CM. Impact of income maintenance on low birthweight: evidence from the Gary experiment. *J Hum Resource* 1979;XIV(4):434-62.
- 24 Bradbury KL, Downs A, eds. *Do housing allowances work?* Washington, DC: Brookings Institution, 1981.
- 25 Friedman J, Weinberg DH, eds. *The great housing experiment*. Beverly Hills, CA: Sage, 1983.
- 26 Bloom HS. *Back to work: testing re-employment services for displaced workers*. Kalamazoo, MI: WE Upjohn Institute for Employment Research, 1990.
- 27 Rossi PH, Berk RA, Leniham KJ. *Money, work and crime: experimental evidence*. New York: Academic Press, 1980.
- 28 Botein B. The Manhattan bail project: its impact on criminology and the criminal law process. *Texas Law Rev* 1965;43:319-31.
- 29 Manning WG, Leibowitz A, Goldberg GA, Rogers WH, Newhouse JP. A controlled trial of the effect of a prepaid group practice on use of services. *N Engl J Med* 1984;310:105-10.
- 30 Gramlich EM, Koshel PP. *Educational performance contracting*. Washington, DC: Brookings Institution, 1975.
- 31 Nathan RP. *Social science and government*. New York: Basic Books, 1988:61.
- 32 Stanley JC. Controlled experimentation in the classroom. *J Exp Educ* 1957;25:195-201.
- 33 Matthews JR. *Quantification and the quest for medical certainty*. Princeton, NJ: Princeton University Press, 1995.
- 34 Chapin FS. The problem of controls in experimental sociology. *J Educ Sociol* 1931;IV:541-51.

(Accepted 1 October 1998)

Words

Nosocomial

Nosocomial, as in *nosocomial infections*, seems to have been around for a long time, but I sense that its usage is increasing, possibly as a result of the problem of antibiotic resistance. When I first saw it some years ago, although I suspected a common origin with *nosology* (which the uninitiated quite reasonably assume to be the missing middle word from oto-laryngology) I had no idea what it meant. Having looked it up, Dorland's *American Illustrated Medical Dictionary* (21st edition) says that it is derived from two Greek words meaning *disease* and *to take care of*, and the *Oxford English Dictionary* (the big one, not the concise) tells me it is derived from an obsolete French word *nosocome* meaning hospital, although it does not appear at all in even the largest of modern French-English dictionaries. I am still uncertain about how it is pronounced; is the third "o" long or short? Whatever the pronunciation, however, as far as I can discover *nosocomial infection* means nothing more nor less than *hospital infection*. Since *nosocomial* is an ugly, affected, and,

to many, an obscure term, which adds nothing to the clarity, precision, or attractiveness of medical discourse or writing, to use it is as daft as using epistaxis for a nosebleed. In support of plain English I suggest that *nosocomial* should be publicly executed and *hospital* put in its place. Any objections?

Irvine Loudon, *medical historian, Wantage, Oxfordshire*

We welcome articles of up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake*, or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk. Permission is needed from the patient or a relative if an identifiable patient is referred to. We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.